

MODIS Document Archive

Presentation to MODIS Technical Team

by Paul Baker and David Herring

October 15, 1992

To : Attachment 2
Date : 10/15/92

MODIS Document Archive

Document Archive Study

Purpose: To define science, technical, and support team requirements for an archive of MODIS-related documents

Objectives:

- 1. Survey team members w/r/t needs and preferences**
- 2. Survey GSFC and other related documents libraries**
- 3. Investigate archiving technologies**
- 4. Determine scope of archive size and users**
- 5. Develop system architecture and cost options**
- 6. Recommend system to meet team needs**

MODIS Document Archive

Survey Results

Documents Libraries

- **EOS and SeaWiFS Libraries hold paper copies of contract deliverable and other documents, maintain PC database of bibliographic data.**
- **Space Station maintains full-text of all SSFPO contract documents on-line electronically. Current architecture requires proprietary software on dedicated workstations for full document viewing.**

Science and Technical Team Members

- **Mailed 29 surveys, received 15 responses**
- **Conducted 15 personal interviews**
- **12 out of 23 Science Team Members responded
(9 survey, 2 personal, 1 both)**

MODIS Document Archive Survey/Interviews

Surveys mailed to:

Abbott, Mark *
Barker, John
Barnes, Bill *
Barton, Ian
Brown, Otis
Carder, Kendall *
Clark, Dennis
Esaias, Wayne
Evans, Robert
Fleig, Al *
Gordon, Howard *
Hall, Dorothy *
Hoge, Frank
Huete, Alfredo
Justice, Chris
Kaufman, Yoram
King, Michael *
Masuoka, Ed *
Menzel, Paul *
Muller, Jan-Peter
Parslow, John
Running, Steven *
Salomonson, Vince
Slater, Philip *
Strahler, Alan *
Stuart, Locke *
Tanre, Didier *
Vanderbilt, Vern
Wan, Zhengming *

Personal interviews with:

Barker, John
Barnes, Bill
Esaias, Wayne
Fleig, Al
Harnden, Joann
Justice, Chris
King, Michael
Lu, Yun-Chi
Masuoka, Ed
McDonald, Ken
Oseroff, Harold
Ramapriyan, H.
Stuart, Locke
Ungar, Stephen
Weber, Richard

* denotes survey response received

MODIS Document Archive Survey Results

I. Documents

For each of the following, choose whether access to the document is:

- a) Essential (Make it "full-text-searchable" on-line)
- b) Somewhat Essential (Make it read-only on-line or make sure I have a copy)
- c) Important, but not essential (I can wait awhile)
- d) Don't throw it away (but, I probably won't look at it)
- e) Not needed in an archive

A. Science & Technical Team Documents

- 1. MODIS Science Team meeting minutes 66% a) or b)
- 2. MODIS Technical Team meeting minutes 66%
- 3. Your Support/Discipline Team meeting minutes 57%
- 4. Team member reports (monthly, quarterly, etc.) 13%
- 5. Team member reports (Semi-annual & annual only) 50%
- 6. Team member-related publications 40%
- 7. Support/Discipline Team technical memoranda 40%
- 8. Support/Discipline Team management information (e.g. budgets) 19%
- 9. Support/Discipline Team handbooks (e.g. calibration handbook) 63%

B. MODIS Contract Deliverables (from CDRL)

- 1. Plans and Review Data Packages (PDR, CDR, etc.) 40% a) or b)
- 2. Math Models and Analyses 32%
- 3. Engineering and Test Reports 38%
- 4. Specifications and Operations 62%
- 5. Procedures 33%
- 6. All MODIS contract documents 7%

C. EOS Project Documents

- 1. EOS Project documents related to MODIS 60% a) or b)
- 2. EOS Project/CDRL documents for other instruments 21%
- 3. All EOS documents 28%

Please list any other MODIS documents, materials (other than data), and instruments you think should be included (with same A, B, C, D, E choices as above)

Letters to and from that are technical, requests for technical info and corresponding answers, MODIS Airborne Simulator, and ASTER documents (all rated a) or b))

II. Level of Interface

1. In what format would you prefer to receive archived documents you request?

- 32% a. Paper (mail it to me)
- 0% b. Paper (fax it to me)
- 0% c. Electronic sneaker-net (send me a disk)
- 42% d. Electronic delivery (e-mail it or leave it for me at anon FTP site)
- 26% e. Electronic access (read online, download all or part thru network)

2. What level of interface sounds like it would best suit your lifestyle?

- 12% a. Just keep me on the distribution list.
6% b. Too busy to get on my computer. Give me a person to call.
29% c. Most of the computer systems I use are command-line based and I would like that style of access (e.g. VT100).
29% d. I use a windows-style interface to most of my computer applications and I would prefer to have that (e.g. X-windows).
24% e. I've used some of the "smart" Internet tools (e.g. Archie, Gopher, WAIS) and I would prefer something like that.

III. Information Tools

For each of the following, estimate whether you would use it:

- a) Regularly (I would use it when I needed any document).
b) Sparingly (I might use it if I couldn't find the right person to ask).
c) Not sure this would help find what I'm looking for.
d) I wouldn't use it, but my staff would.
e) Wouldn't use it.

	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>	<u>e</u>
1. Card Catalog	3	4	2	2	4
2. List of Documents					
by title	3	6	3	1	1
by author	5	6	2	1	1
by subject	6	6	1	1	1
by date	4	4	3	1	2
3. Bibliographic Database					
search by title, author, subject, or date	10	3	1	1	0
4. Full-text Search					
some documents	6	7	1	0	0
all documents in system	3	3	2	2	2
5. Other Reference Tools					
glossary	4	6	2	1	2
list of acronyms	7	4	2	0	2
6. (deleted)					

Are there any tools you think should be added to the list (please rate as above)?

Loss-less data compression for scanned images, conversion tools for text documents to and from UNIX environment

MODIS Document Archive

Size of Potential Document Archive (including MCST)

- **Current Archive ~ 63,000 pages**
- **Potential Archive (2003) ~ 1,271,000 pages**
- **Computer Storage Required:**

300 dpi scan resolution @ 20:1 compression = 15,000 pgs per 1 Gb

MODIS storage requirement = 4.5 Gb now, 85 Gb in 10 years

Scope and Size of User Population

- **23 ST Members and ~40 TT/Supp users**
- **> 50% of total users located at Goddard**
- **expecting regular usage (> twice a week) from ~ 5 -10 users**

MODIS Archive Storage Needs

Document Types	No. pgs	Proj. pgs (10 years)	Cur Storage Req. (15K pgs/Gb/1 file cab)	10-Year Storage Req. (15K pgs/Gb/1 file cab)
MAST DOCS				
ST Meeting Minutes[1]	3,500.00	43,500.00	0.23	2.90
TT Meeting Minutes[2]	600.00	3,000.00	0.04	0.20
M,Q,S,A Reports[3]	550.00	8,850.00	0.04	0.59
Miscellaneous[4]	2,000.00	21,000.00	0.13	1.40
TOTAL	6,650.00	76,350.00	0.44	5.09
MCST DOCS[5]				
SBRC Docs	30,000.00	428,000.00	2.00	28.53
GE Docs	1,500.00	250,000.00	0.10	16.67
MCST Docs	6,000.00	96,000.00	0.40	6.40
Ancillary IP/OP	15,000.00	300,000.00	1.00	20.00
TOTAL	52,500.00	1,074,000.00	3.50	71.60
SDST DOCS[6]				
SDST Meeting Minutes	200.00	2,000.00	0.01	0.13
SDST Presentations	1,800.00	27,000.00	0.12	1.80
Technical Docs	1,800.00	90,000.00	0.12	6.00
Technical Memos	100.00	1,500.00	0.01	0.10
TOTAL	3,800.00	120,500.00	0.25	8.03
GRAND TOTAL	62,950.00	1,270,850.00	4.20	84.72

[1] Meeting frequency may increase to 4 per year

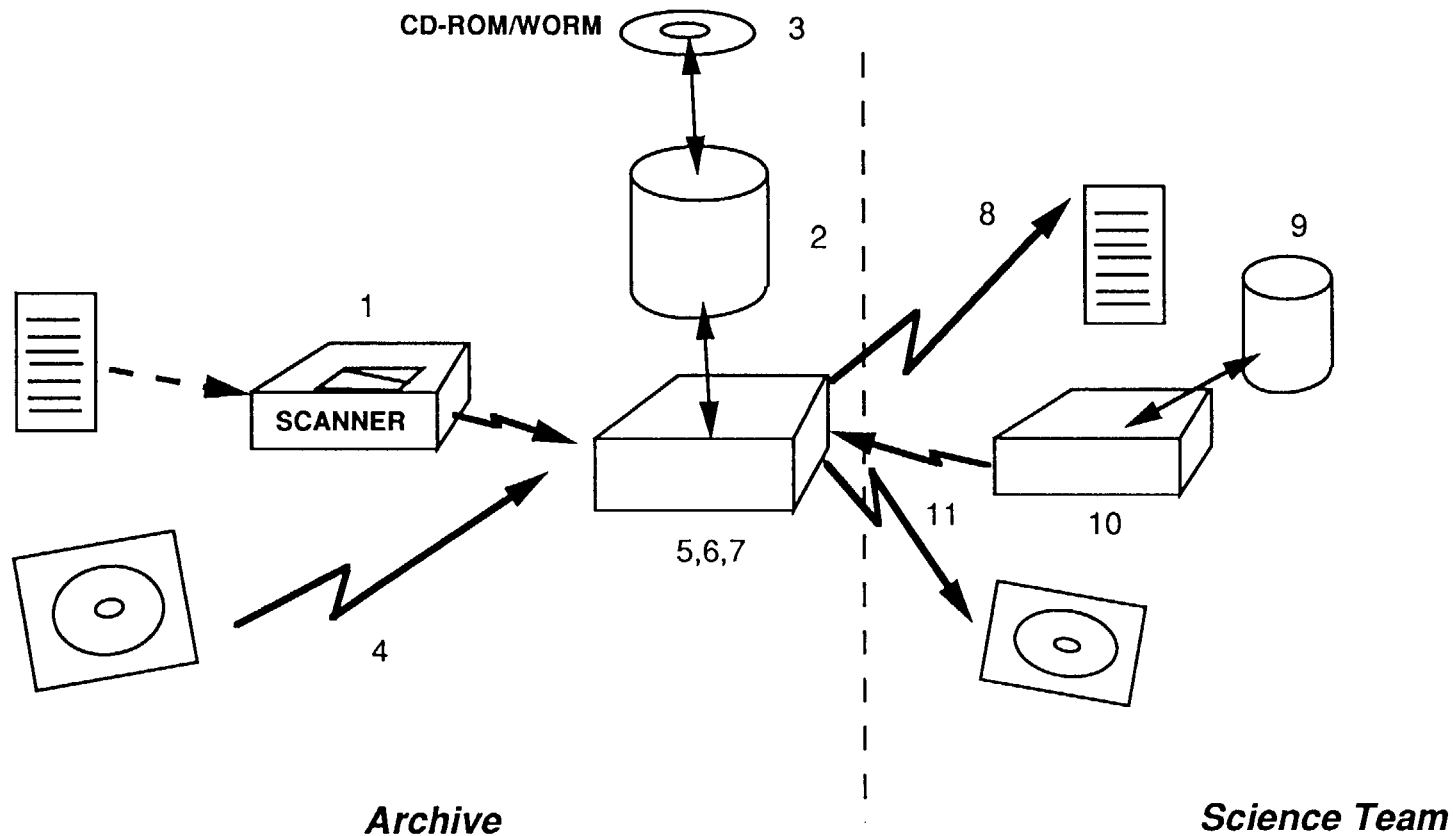
[2] Meeting frequency recently increased to once per week

[3] Assumes the number of Science Team Members is constant

[4] There is some redundancy among SDST, MCST, and MAST archives

[5] MCST expects the rate of input and output of documents to increase greatly

[6] The number of SDST personnel may increase by up to 5 times. The number of documents input and output will increase by 3 or 4 times



1. Documents scanner
2. Mass storage access
3. Storage media
4. Electronic submission
5. Indexing software
6. OCR software
7. Image compression

8. Print capability
9. Bibliographic database
10. Reference tools
11. Electronic access/delivery

MODIS Document Archive System

MODIS Document Archive

Required Features

- **Open system (UNIX, TCP/IP, X-windows)**
- **Full-text search and retrieval including fuzzy search to compensate for OCR errors**
- **Makes use of existing MODIS ST/TT hardware where possible**
- **Supports Mac and PC-based Clients**
- **Auto-indexes text and images**
- **Supports remote scanning and printing, batch OCR, and image compression**
- **Requires minimal "overhead" for index storage**

MDAS Features Matrix

Features	MD Mars	Optix	EFS/PixTex
Document Ingest			
Remote Scanning Control	x	x	x
Open Scanner Interface	x	x	x
Image Compression	x	x	x
Optical Character Recognition	x	x	x
Batch OCR Processing	x	x	x
Image Server			
Supports Mac Clients	x	x	x
Supports PC Clients	x		x
Supports VT-100		x	x
Supports X-windows			x
Multi-user,-tasking O/S		x	x
Supports SQL Databases		x	x
Supports Forms Drop-outs		x	
Supports zooming, moving	x	x	x
Network			
Supports TCP/IP	x	x	x
Supports Novell			x
Supports Appletalk	x	x	
Supports DECNet			x
Supports Ethernet	x	x	x
Index and Retrieval			
Supports Auto-indexing		x	x
Doc. Index < 50% of ASCII req.			x
User-configurable fields/labels		x	x
Fuzzy search capability	x		x
Boolean search	x	x	x
Relevance ranking of hits	x		x
Printing, Export			
Supports remote printing	x	x	x
Supports Mac and other WP	x	x	x
Print ASCII & images	x	x	x

MODIS Document Archive

Recommendation -- MODIS Electronic Document Archive

First year hardware/software & operator costs:

	<u>Total Cost</u>	<u>Trial Cost</u>
Excalibur's EFS/PixTex (bundled w/ OCR, indexing, compression and fuzzy full-text retrieval S/W)	\$52K	10 clients-\$40K
Sun SPARCServer 2 w/ Tape Back-up	\$25K	borrowed
Document Scanner	\$7K	\$7K
47 Gb Optical Storage	\$50K	\$8K for 2 Gb
Operator/ Training	\$43K	\$43K
	-----	-----
First Year Total	\$177K	\$98K
Total estimated 5 yr. cost	\$308.2K	

MODIS Docs. Arch. 5 Yr. Budget

ITEMS	Year 1	Year 2	Year 3	Year 4	Year 5	TOTAL
Ingest						
Scanner	7,000	0	0	0	0	7,000
Server w/Tape B/Up	25,000	0	0	0	0	25,000
Optical Storage	50,000	0	0	0	0	50,000
Operator	40,000	20,000	20,000	20,000	20,000	120,000
Maint./Upgrades	0	8,200	8,200	8,200	8,200	32,800
Training	3,000	0	0	0	0	3,000
Integration	3,000	0	0	0	0	3,000
Retrieval						
Workstation Clients	18,000	0	0	0	0	18,000
System Software	28,000	0	0	0	0	28,000
Training	3,000	0	0	0	0	3,000
Maint./Upgrades	0	4,600	4,600	4,600	4,600	18,400
TOTAL	177,000	32,800	32,800	32,800	32,800	308,200
5 yr. max cost w/operator= 308,200						